Linear Algebra Notes

Nikhil Srivastava

February 9, 2015

Scalars are lowercase, matrices are uppercase, and vectors are lowercase **bold**. All vectors are column vectors (i.e., a vector in $\mathbb{R}^n$ is an $n \times 1$ matrix), unless transposed.

1 Column Picture of Matrix Vector Multiplication

Suppose A is an $m \times n$ matrix with rows $\mathbf{r}_1^T, \ldots, \mathbf{r}_m^T$ and columns $\mathbf{c}_1, \ldots, \mathbf{c}_n$. In high school, we are taught to think of the matrix-vector product $A\mathbf{x}$ as taking dot products of $\mathbf{x}$ with rows of A:

$$A\mathbf{x} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \cdots \\ \mathbf{r}_m^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{r}_1^T\mathbf{x} \\ \mathbf{r}_2^T\mathbf{x} \\ \cdots \\ \mathbf{r}_m^T\mathbf{x} \end{bmatrix} = \begin{bmatrix} (\mathbf{r}_1 \cdot \mathbf{x}) \\ (\mathbf{r}_2 \cdot \mathbf{x}) \\ \cdots \\ (\mathbf{r}_m \cdot \mathbf{x}) \end{bmatrix}$$

This makes sense if we regard matrices as essentially a convenient notation to represent linear equations, since each linear equation naturally gives a dot product.

A different perspective is to view $A\mathbf{x}$ as taking a linear combination of the *columns* $\mathbf{c}_1, \ldots, \mathbf{c}_n$ of A, with coefficients equal to the entries of $\mathbf{x}$:

$$A\mathbf{x} = [\mathbf{c}_1|\mathbf{c}_2|\ldots|\mathbf{c}_n] \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \ldots + x_n\mathbf{c}_n.$$

This view makes it transparent that matrices can be used to represent linear transformations with respect to a pair of bases. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis of $\mathbb{R}^n$. Then we can write any $\mathbf{x} \in \mathbb{R}^n$ as

$$\mathbf{x} = x_1\mathbf{e}_1 + \ldots + x_n\mathbf{e}_n$$

for some coefficients x_i; indeed, this is what we mean when we identify $\mathbf{x}$ with its standard coordinate vector:

$$[\mathbf{x}] = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}.$$

Since T is linear, it is completely determined by its values on the basis:

$$T(\mathbf{x}) = T(x_1\mathbf{e}_1 + \ldots + x_n\mathbf{e}_n) = x_1T(\mathbf{e}_1) + \ldots x_nT(\mathbf{e}_n).$$

Since these vectors are equal, their coordinate vectors in the standard basis of the "output" vector space $\mathbb{R}^m$ must also be equal:

$$[T(\mathbf{x})] = [T(x_1\mathbf{e}_1 + \ldots + x_n\mathbf{e}_n)] = x_1[T(\mathbf{e}_1)] + \ldots x_n[T(\mathbf{e}_n)].$$

But since matrix-vector multiplication is the same thing as taking linear combinations, we may write this as

$$[T(\mathbf{x})] = \big[[T(\mathbf{e}_1)]|[T(\mathbf{e}_2)]|\ldots|[T(\mathbf{e}_n)]\big] \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix},$$

where $[\mathbf{v}]$ denotes the standard coordinate vector of $\mathbf{v}$. Note that anything that appears inside a matrix must be some sort of coordinate vector: it does not make sense to put an abstract vector inside a matrix whose entries are numbers. However, as before, we will identify $\mathbf{v}$ with its standard coordinate vector $[\mathbf{v}]$ and drop the brackets when we are working in the standard basis. With this identification, we can write:

$$T(\mathbf{x}) = \big[T(\mathbf{e}_1)|T(\mathbf{e}_2)|\ldots|T(\mathbf{e}_n)\big] \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix}.$$

The matrix above is called the standard matrix of T, and is denoted by $[T]$. It is a complete description of T, with respect to the standard basis. One of the remarkable things about linear transformations is that they have such compact descriptions — this is not at all true of arbitrary functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Remark. The column view of matrix-vector multiplication also explains why matrix-matrix multiplication is defined the way it is (which usually seems mysterious the first time you see it). It is the unique definition for which

$$[S \circ T] = [S][T].$$

where $(S \circ T)(\mathbf{v}) = S(T(\mathbf{v}))$ denotes the composition of two linear transformations S and T. More concretely, for any matrix A and column vectors $\mathbf{c}_1 \ldots, \mathbf{c}_n$, it is the unique definition for which

$$A\left[\mathbf{c}_1|\mathbf{c}_2|\ldots|\mathbf{c}_n\right] = \left[A\mathbf{c}_1|A\mathbf{c}_2|\ldots|A\mathbf{c}_n\right].$$

2 Change of Basis

We know we can write any $\mathbf{x} \in \mathbb{R}^n$ as a linear combination of the standard basis vectors, for some coefficients $x_1, \ldots, x_n$:

$$\mathbf{x} = x_1\mathbf{e}_1 + \ldots + x_n\mathbf{e}_n.$$

In some situations, it is useful to write $\mathbf{x}$ in another basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ (we will see many such situations in the course). By definition, since the $\mathbf{b}_i$'s are a basis, there must be unique coefficients $x'_1, \ldots, x'_n$ such that

$$\mathbf{x} = x'_1\mathbf{b}_1 + x'_2\mathbf{b}_2 + \ldots + x'_n\mathbf{b}_n.$$

Passing to this representation of $\mathbf{x}$ is what I called "implicit" change of basis in class. By implicit, I meant that we chose to decompose $\mathbf{x}$ in terms of its coefficients in the $\mathbf{b}_i$, but we didn't give an explicit method for computing what those coefficients are.

To find the coefficients, it helps to write down the latter linear combination in matrix notation:

$$[\mathbf{x}] = x'_1[\mathbf{b}_1] + \ldots + x'_n[\mathbf{b}_n] = [\mathbf{b}_1|\ldots|\mathbf{b}_n]\begin{bmatrix} x'_1 \\ \ldots \\ x'_n \end{bmatrix} = B[\mathbf{x}]_B,$$

where B is the matrix with columns $\mathbf{b}_1, \ldots, \mathbf{b}_n$ and $[\mathbf{x}]_B$ denotes the coordinate vector of $\mathbf{x}$ in the B basis. This yields the fundamental relationship

$$[\mathbf{x}] = B[\mathbf{x}]_B,$$

which since B is invertible tells us how to find $[\mathbf{x}]_B$ from $[\mathbf{x}]$:

$$[\mathbf{x}]_B = B^{-1}[\mathbf{x}].$$

Thus, changing basis is equivalent to solving linear equations. Note that this also shows that the columns of B^{-1} are $[\mathbf{e}_1]_B, \ldots, [\mathbf{e}_n]_B$, the standard basis vectors in the B basis.

Once we know how to change basis for vectors, it is easy to do it for linear transformations and matrices. For simplicity, we will only consider linear operators (linear transformations from a vector space to itself), which will allow us to keep track of just one basis rather than two.

Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator and $[T]$ is its standard matrix, i.e.

$$[T][\mathbf{x}] = [T(\mathbf{x})].$$

By definition, the matrix $[T]_B$ of T in the B basis must satisfy:

$$[T]_B[\mathbf{x}]_B = [T(\mathbf{x})]_B.$$

Plugging in the relationship between $[\mathbf{x}]_B$ and $[\mathbf{x}]$ which we derived above, we get

$$[T]_B B^{-1}[\mathbf{x}] = B^{-1}[T(\mathbf{x})],$$

which since B is invertible is equivalent to

$$B[T]_B B^{-1}[\mathbf{x}] = [T(\mathbf{x})].$$

Thus, we must have

$$B[T]_B B^{-1} = [T],$$

or equivalently

$$[T]_B = B^{-1}[T]B,$$

which are the explicit formulas for change of basis of matrices/linear transformations.

3 Diagonalization

The standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of $\mathbb{R}^n$ is completely arbitrary, and as such it is just a convention invented by humans so that they can start writing vectors in *some* basis. But linear transformations that occur in nature and elsewhere often come equipped with a much better basis, given by their eigenvectors.

Let $T : V \to V$ be a linear operator (i.e., a linear transformation from a vector space to itself). If $\mathbf{v} \neq 0$ is a nonzero vector and $T(\mathbf{v}) = \lambda\mathbf{v}$ for some scalar λ (which may be complex), then $\mathbf{v}$ is called an *eigenvector* of T and λ is called an *eigenvalue* of $\mathbf{v}$.

There is an analogous definition for square matrices A, in which we ask that $A\mathbf{v} = \lambda\mathbf{v}$. Note that $T(\mathbf{v}) = \lambda\mathbf{v}$ if and only if $[T][\mathbf{v}] = \lambda[\mathbf{v}]$, so in particular the operator T and the matrix $[T]$ have the same eigenvalues. This fact holds in every basis (see HW3 question 6), so eigenvalues are intrinsic to operators and do not depend on the choice of basis used to write the matrix.

The eigenvalues of a matrix may be computed by solving the characteristic equation $\det(\lambda I - A) = 0$. Since this is a polynomial of degree n for an $n \times n$ matrix, the fundamental theorem of algebra tells us that it must have n roots, whence every $n \times n$ matrix must have n eigenvalues. Once the eigenvalues are known, the corresponding eigenvectors can be obtained by solving systems of linear equations $(\lambda I - A)\mathbf{v} = 0$. See your Math 54 text for more information on how to compute these things.

In any case, something wonderful happens when we have an operator/matrix with a *basis* of linearly independent eigenvectors, sometimes called an eigenbasis. For instance, let $T : \mathbb{R}^n \to \mathbb{R}^n$ be such an operator. Then, we have

$$T(\mathbf{b}_1) = \lambda_1\mathbf{b}_1, T(\mathbf{b}_2) = \lambda_2\mathbf{b}_2, \ldots, T(\mathbf{b}_n) = \lambda_n\mathbf{b}_n,$$

for linearly independent $\mathbf{b}_1, \ldots, \mathbf{b}_n$. Thus, we may write an arbitrary $\mathbf{x} \in \mathbb{R}^n$ in this basis:

$$\mathbf{x} = x'_1\mathbf{b}_1 + x'_2\mathbf{b}_2 + \ldots + x'_n\mathbf{b}_n,$$

and then

$$\begin{aligned} T(\mathbf{x}) &= T(x'_1\mathbf{b}_1 + x'_2\mathbf{b}_2 + \ldots + x'_n\mathbf{b}_n) \\ &= x'_1T(\mathbf{b}_1) + x'_2T(\mathbf{b}_2) + \ldots + x'_nT(\mathbf{b}_n) \quad \text{by linearity of } T \\ &= x'_1\lambda_1\mathbf{b}_1 + x'_2\lambda_2\mathbf{b}_2 + \ldots + x'_n\lambda_n\mathbf{b}_n. \end{aligned}$$

So, applying the transformation T is tantamount to multiplying each coefficient x'_i by λ_i. In particular T acts on each coordinate completely independently by *scalar* multiplication, and there are no interactions between the coordinates. This is as about as simple as a linear transformation can be.

If we write down the matrix of T in the basis B consisting of its eigenvectors, we find that $[T]_B$ is a diagonal matrix with the eigenvalues $\lambda_1, \ldots, \lambda_n$ on the diagonal. Appealing to the change of basis formula we derived in the previous section, this means that

$$[T] = B[T]_B B^{-1}.$$

In general, this can be done for any square matrix A, since every A is equal to $[T]$ for the linear transformation $T(\mathbf{x}) = A\mathbf{x}$. Using the letter D to denote the diagonal matrix of eigenvalues, this gives

$$A = B^{-1}DB.$$

Factorizing a matrix in this way is called *diagonalization*, and a matrix which can be diagonalized (i.e., one with a basis of eigenvectors) is called *diagonalizable*. Not all matrices are diagonalizable, but the vast majority of them are.

4 Coupled Oscillator Example

Here is an example of diagonalization in action. Suppose we have two unit masses connected by springs with spring constant k as in Figure 12.1 of the book. Let $x_1(t)$ and $x_2(t)$ be the positions of the springs at time t. Then, subject to initial positions and velocities $x_1(0), x_2(0), \dot{x_1}(0), \dot{x_2}(0)$, the system is governed by the coupled differential equations:

$$\frac{\partial^2 x_1(t)}{\partial^2 t} = -kx_1(t) - k(x_1(t) - x_2(t)),$$

$$\frac{\partial^2 x_2(t)}{\partial^2 t} = -kx_2(t) - k(x_2(t) - x_1(t)).$$

It is not immediately obvious how to solve this because each partial derivatives depends on both of the variables. We will now show how to reduce it to the diagonal case.

We can write this as a single differential equation in the vector-valued function

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

as follows:

$$\frac{\partial^2 \mathbf{x}(t)}{\partial t^2} = A\mathbf{x}(t),$$

where

$$A = \begin{bmatrix} -2k & k \\ k & -2k \end{bmatrix}$$

It turns out that A has eigenvalues $-k$ and $-3k$ and corresponding eigenvectors

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{b}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

We will now show how this simplifies the problem substantially, first with an "implicit" change of basis, and then with an "explicit" matrix factorization.

Implicit Change of Basis. For every value of t, $\mathbf{x}(t)$ is a vector in $\mathbb{R}^2$, so we can write it as some linear combination of the basis vectors $\{\mathbf{b}_1, \mathbf{b}_2\}$:

$$\mathbf{x}(t) = a_1(t)\mathbf{b}_1 + a_2(t)\mathbf{b}_2, \qquad (*)$$

where the coefficients $a_1(t)$ and $a_2(t)$ depend on t. Substituting this into our equations, we have

$$\begin{aligned} \frac{\partial^2 a_1(t)}{\partial t^2}\mathbf{b}_1 + \frac{\partial^2 a_2(t)}{\partial t^2}\mathbf{b}_2 &= \frac{\partial^2}{\partial t^2}\left(a_1(t)\mathbf{b}_1 + a_2(t)\mathbf{b}_2\right) \\ &= A(a_1(t)\mathbf{b}_1 + a_2(t)\mathbf{b}_2) \\ &= a_1(t)A\mathbf{b}_1 + a_2(t)A\mathbf{b}_2 \\ &= a_1(t)\lambda_1\mathbf{b}_1 + a_2(t)\lambda_2\mathbf{b}_2. \end{aligned}$$

Since $\mathbf{b}_1, \mathbf{b}_2$ are linearly independent, their coefficients on both sides of this equality must be equal. Equating coefficients, we obtain the decoupled scalar differential equations:

$$\begin{aligned} \frac{\partial^2 a_1(t)}{\partial t^2} &= \lambda_1 a_1(t), \\ \frac{\partial^2 a_2(t)}{\partial t^2} &= \lambda_2 a_2(t), \end{aligned}$$

For which we can easily find the general solutions:

$$a_1(t) = a_1(0)\cos(\sqrt{-\lambda_1}t) + \frac{\dot{a_1}(0)}{\sqrt{-\lambda_1}}\sin(\sqrt{-\lambda_1}t),$$

and

$$a_2(t) = a_2(0)\cos(\sqrt{-\lambda_2}t) + \frac{\dot{a_2}(0)}{\sqrt{\lambda_2}}\sin(\sqrt{-\lambda_2}t).$$

The initial conditions we had in class were $x_1(0) = 1$ and $x_2(0) = \dot{x_1}(0) = \dot{x_2}(0) = 0$. We again use the equation $(*)$ to translate this into initial conditions in $a_1(t), a_2(t)$:

$$\mathbf{x}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = (1/2)\mathbf{b}_1 + (1/2)\mathbf{b}_2,$$

so $a_1(0) = a_2(0) = 1/2$. We also have $\dot{a_1}(0) = \dot{a_2}(0) = 0$. Thus, the general solution is given by

$$\mathbf{a}(t) = \begin{bmatrix} a_1(t) \\ a_2(t) \end{bmatrix} = 1/2\begin{bmatrix} \cos(\sqrt{k}t) \\ \cos(\sqrt{3k}t) \end{bmatrix},$$

which in terms of $\mathbf{x}$ is just

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \cos(\sqrt{k}t) + \cos(\sqrt{3k}t) \\ cos(\sqrt{k}t) - \cos(\sqrt{3k}t) \end{bmatrix}$$

Explicit Matrix Notation. Some people understand things better if they are written in terms of explicit matrix factorizations. If we diagonalize $A = BDB^{-1}$, we can rewrite our equation as

$$\frac{\partial^2}{\partial t^2}\mathbf{x}(t) = BDB^{-1}\mathbf{x}(t).$$

Defining

$$\mathbf{a}(t) = B^{-1}\mathbf{x}(t) \qquad \text{(this is the same as } (*)\text{, in matrix notation),}$$

this becomes

$$\frac{\partial^2}{\partial t^2}B\mathbf{a}(t) = BD\mathbf{a}(t).$$

Since B is a fixed matrix that does not depend on t, it commutes with the partial derivative and we have

$$B\frac{\partial^2}{\partial t^2}\mathbf{a}(t) = BD\mathbf{a}(t).$$

Multiplying both sides by B^{-1} gives

$$\frac{\partial^2}{\partial t^2}\mathbf{a}(t) = D\mathbf{a}(t),$$

which is the same diagonal system we solved above.

The eigenvectors of A are called the *normal modes* of the system.

5 Symmetric and Orthogonal Matrices

Not all matrices are diagonlizable, but there are some very important classes that are. A real matrix A is called symmetric if $A = A^T$. To state the main theorem about diagonalizing symmetric matrices, we will need a definition.

Definition. A collection of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is called orthonormal if they are pairwise orthogonal, i.e.,

$$(\mathbf{v}_i \cdot \mathbf{v}_j) = 0 \quad \text{for } i \neq j$$

and they are unit vectors:

$$(\mathbf{v}_i \cdot \mathbf{v}_i) = 1.$$

In matrix notation, an orthonormal set of vectors has the property that

$$V^T V = I,$$

where $V = [\mathbf{v}_1|\ldots|\mathbf{v}_n]$ is a matrix with the $\mathbf{v}_i$ as columns; such a matrix is called an orthogonal matrix.

This last identity implies that

$$V^{-1} = V^T,$$

which reveals one of the very desirable properties of orthonormal bases: for any $\mathbf{x}$, the change of basis is simply

$$[\mathbf{x}]_V = V^{-1}[\mathbf{x}] = V^T[\mathbf{x}] = \begin{bmatrix} (\mathbf{v}_1 \cdot \mathbf{x}) \\ \ldots \\ (\mathbf{v}_n \cdot \mathbf{x}) \end{bmatrix}.$$

That is, the coefficients are given by *dot products*

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{x} \cdot \mathbf{v}_2)\mathbf{v}_2 + \ldots + (\mathbf{x} \cdot \mathbf{v}_n)\mathbf{v}_n,$$

which are much easier to calculate than solving linear equations as in Section 2

Theorem. If A is symmetric, then it is diagonalizable. Moreover, all of its eigenvalues are real and it has an eigenbasis of *orthonormal* eigenvectors. Thus, $A = VDV^T$ for some orthogonal V and diagonal D.

6 Complex Inner Products, Hermitian Matrices, and Unitary Matrices

There is an important class of complex matrices which are also diagonalizable with orthogonal eigenvectors. However, the notion of orthogonality (which is a geometric notion induced by the dot product) is different for complex vectors.

To see that the usual real dot product is deficient in the complex case, consider that

$$\begin{bmatrix} 1 \\ i \end{bmatrix} \cdot \begin{bmatrix} 1 \\ i \end{bmatrix} = 1^2 + i^2 = 0,$$

so we have a nonzero vector which is orthogonal to itself. This makes no sense geometrically. It turns out that there is a way to *redefine* the dot product which recovers all the nice geometric properties that we have in the real case:

$$\langle \mathbf{x} | \mathbf{y} \rangle = x_1^* y_1 + x_2^* y_2 + \ldots + x_n^* y_n.$$

This is the same as the real dot product, except we take the complex conjugate (denoted by $*$) of the first vector $\mathbf{x}$. Note that if both $\mathbf{x}$ and $\mathbf{y}$ are real then this doesn't change anything. We will refer to this as an "inner product" to distinguish it from the real dot product.

With the inner product in hand, we say that a set of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ in $\mathbb{C}^n$ is orthonormal if

$$\langle \mathbf{u}_i | \mathbf{u}_j \rangle = 0 \quad \text{for } i \neq j$$

and

$$\langle \mathbf{u}_i | \mathbf{u}_i \rangle = 1.$$

A matrix with complex orthonormal columns is called *unitary.* You should check that such a matrix satisfies

$$(U^*)^T U = I.$$

This looks a ilttle weird, but it's only because the usual transpose isn't the correct notion for complex matrices. The right generalization is actually the *conjugate transpose*

$$U^\dagger := (U^T)^*,$$

pronounced U "dagger". In this notation, a unitary matrix is just one which satisfies

$$U^\dagger = U^{-1}.$$

Again, computing coefficients in such a basis is very easy, and amounts to finding inner products:

$$\mathbf{x} = \langle \mathbf{u}_1|\mathbf{x}\rangle \mathbf{u}_1 + \ldots + \langle \mathbf{u}_n|\mathbf{x}\rangle.$$

The correct generalization of real symmetric matrices to the complex case is the class of *Hermitian* matrices. A matrix A is called Hermitian if $A = A^\dagger$.

Theorem. If A is Hermitian, then it is diagonalizable. Moreover, all of its eigenvalues are real and it has an eigenbasis of *orthonormal* eigenvectors. Thus, $A = UDU^\dagger$ for some unitary U and diagonal D.

7 Inner Product Spaces

It turns out that the notion of an inner product can be made to work in even more general settings than $\mathbb{C}^n$. The motivation for doing this is that the inner product is the device that allows us to define lengths and angles in Euclidean geometry.

Definition. A function $\langle\cdot|\cdot\rangle : V \times V \to \mathbb{C}$ is called an inner product if it satisfies the following properties:

1. $\langle \mathbf{x}|\mathbf{x}\rangle \geq 0$ for all $\mathbf{x}$, with equality iff $\mathbf{x} = 0$.
2. $\langle a\mathbf{x}_1 + b\mathbf{x}_2|\mathbf{y}\rangle = a^*\langle \mathbf{x}_1|\mathbf{y}\rangle + b^*\langle \mathbf{x}_2|\mathbf{y}\rangle$ for all $a, b \in \mathbb{C}$ and $\mathbf{x}, \mathbf{y} \in V$. This is called "conjugate-linearity" in the first coordinate.
3. $\langle \mathbf{x}|\mathbf{y}\rangle^* = \langle \mathbf{y}|\mathbf{x}\rangle$. In particular, this implies linearity in the second coordinate.

An inner product induces a *norm* $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}|\mathbf{x}\rangle}$, whose intended interpretation is that it is the length of the vector $\mathbf{x}$.

It turns out that these three properties are all we need to guarantee that the familiar theorems of Euclidean geometry continue to hold with the new inner product. This is incredibly powerful; later in the course we will define inner products on vector spaces of functions, and it will allows us to "visualize" them in the familiar ways that we visualize Euclidean space.

Here are three important properties shared by all inner products:

1. $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. (triangle inequality)
2. If $\langle \mathbf{x}|\mathbf{y}\rangle = 0$ then $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$. (Pythagoras Theorem)
3. $|\langle \mathbf{x}|\mathbf{y}\rangle| \leq \|\mathbf{x}\|\|\mathbf{y}\|$. (Cauchy-Schwartz Inequality).

Item (2) is a simple exercise, and item (1) can be easily derived from item (3). Here is the proof of item (3): first observe that it is equivalent to show that

$$\left|\left\langle \frac{\mathbf{x}}{\|\mathbf{x}\|} \middle| \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\rangle\right| = \frac{|\langle \mathbf{x}|\mathbf{y}\rangle|}{\|\mathbf{x}\|\|\mathbf{y}\|} \leq 1,$$

where the first equality is because of linearity in both coordinates with respect to real scalars. So it suffices to show that

$$|\langle \mathbf{x}|\mathbf{y}\rangle| \leq 1$$

for all unit vectors $\mathbf{x}, \mathbf{y}$. We now compute

$$\langle \mathbf{x}-\mathbf{y}|\mathbf{x}-\mathbf{y}\rangle = \langle \mathbf{x}|\mathbf{x}\rangle + \langle \mathbf{y}|\mathbf{y}\rangle - \langle \mathbf{y}|\mathbf{x}\rangle - \langle \mathbf{x}|\mathbf{y}\rangle.$$